DOI:10.11931/guihaia.gxzw201906012

降香黄檀叶绿体基因组密码子偏好性分析

原晓龙，李云琴，张劲峰，王毅*

(云南省林业和草原科学院，云南省森林植物培育与开发利用重点实验室，国家林业和草原局云南珍稀濒特森林植物保护和繁育重点实验室，昆明 650201)

摘要：为了解降香黄檀叶绿体基因组密码子使用模式，本研究利用 Codon W 1.4.2 和在线软件 CUSP 对降香黄檀叶绿体基因组中的 52 条基因编码序列密码子进行中性绘图、ENC-plot 和 PR2-plot 分析。结果表明：降香黄檀叶绿体基因组密码子的 3 个位置上 GC 含量依次为 GC_1 (46.01%) >GC_2 (38.98%) >GC_3 (27.80%)；有效密码子数(ENC)范围为 37.66 ~ 54.43，及 ENC 值>45 的有 37 个；RSCU >1 的密码子有 29 个，其中 16 个以 U 结尾、12 个以 A 结尾；这些说明其偏好以 AT 结尾，且偏性较弱。中性绘图分析显示 GC_{12} 与 GC_3 的相关系数 0.250，相关性不显著，回归系数为 0.394；ENC-plot 分析显示 ENC 比值位于-0.05 ~ 0.15 区间的基因有 39 个；且 PR2-plot 分析在碱基的使用频率方面，U > A、G > C；说明降香黄檀叶绿体基因组密码子偏好性主要受选择的影响；19 个密码子被确定最优密码子。本研究为降香黄檀叶绿体基因工程、遗传多样性分析等研究提供了科学参考依据。

关键词：降香黄檀, 叶绿体基因组, 密码子偏好性, 选择

中图分类号：Q755 **文献标识码：**A

Analysis of codon usage bias in the chloroplast genome of *Dalbergia odorifera*

Yuan Xiaolong, Li Yunqin, Zhang Jinfeng, Wang Yi*

(*Yunnan Provincial Key Laboratory of Cultivation and Exploitation of Forest Plants*, *Conservation of Rare, Endangered & Endemic Forest Plants*/ *Public Key Laboratory of National Forestry and Grassland Administration*, *Yunnan Academy of Forestry and Grassland Science*, Kunming 650201)

基金项目：国家自然科学基金(31860177)；云南省林业科学院热带亚热带珍贵用材树种研发省创新团队(2017HC024)[Supported by the National Natural Science Foundation of China(31860177); Provincial Innovation Group of R & D of Precious Tropical and Subtropical Timber Species in Yunnan Academy of Forestry (2017HC024)]。

作者简介：原晓龙(1986-)，男，陕西蒲城人，硕士，助理研究员，主要从事林木分子生物学研究，(E-mail)xiaolony@126.com。

***通信作者：**王毅，博士，助理研究员，主要从事植物学和分子生物学研究，(E-mail)22825818@qq.com。

Abstract: To comprehend the codon usage pattern of chloroplast genome in *Dalbergia odorifera*, the 52 coding DNA sequences was analyzed to obtain to the results of neutrality plot, ENC-plot and PR2-plot analysis using Codon W 1.4.2 and online software CUSP in the present study. The results showed that the GC content in the three positions of codons from the chloroplast genome of *D. odorifera* was GC_1 (46.01%) >GC_2 (38.98%) >GC_3 (27.80%) successively. The range of effective number codon was from 37.66 to 54.43, and there were 37 genes when ENC value was greater than 45. There were 29 genes when RSCU value was greater than 1, including the ending of 16 genes were U and 12 genes were A. These suggested that the codons preferred ends with AT, and had a weak bias. The neutrality plot showed that there was no significant correlation between GC_3 and GC_{12}, the correlation coefficient was 0.250, and the regression coefficient was 0.394; ENC-plot analysis revealed that there were 39 genes which ENC ratio located in the section from -0.05 to 0.15; PR2-plot analysis showed that U>A and G>C in the base usage frequency. These all illustrated that the codon usage bias in the chloroplast genome of *D. odorifera* was mainly affected by mutation; 19 codons were identified as the optimal codon. The present study could be useful in the chloroplast genetic engineering and genetic diversity analysis of *D. odorifera*.

Keywords: *Dalbergia odorifera*, chloroplast genome, codon usage bias, selection

在生物体传递遗传信息的过程中，作为联结核酸和蛋白质的密码子扮演着重要的角色(Zhang et al., 2019)，构成基因组的 4 种核苷酸可形成 64 种密码子，各密码子与氨基酸相对应，除甲硫氨酸和色氨酸外，其余 18 种氨基酸均有 2 ~ 6 个密码子，这些编码同一氨基酸的不同密码子被称为同义密码子(synonymous codon)(Duret, 2002)；在翻译过程中，每个氨基酸相对应同义密码子的使用频率存在差异，即有的同义密码子使用频率高于其他同义密码子，这种现象被称为密码子偏好性(codon usage bias)(Romero et al., 2000)。密码子偏好性广泛存在于不同生物中，是物种在长期进化过程中受环境选择、碱基突变、基因漂变等因素共同作用，还受到基因组大小、tRNA 丰度和基因表达水平等的影响(Romero et al., 2000; Duret, 2000; Xu et al., 2011)。密码子偏好性通过对基因翻译准确性和效率的调节影响基因的表达水平(邢朝斌等, 2013)，叶绿体具有基因组小、基因拷贝数多等特点(Wright, 1990; 牛元等, 2018)；根据最优密码子设计叶绿体基因表达载体，可迅速提高叶绿体基因组中基因表达量，利用目前已知的密码子使用模式推断未知基因的表达，或预测某些未知基因的功能(Wu et al, 2007)；

同时亲缘关系较近的物种具相近的密码子使用模式(杨国锋等, 2015)，因此研究叶绿体基因组密码子的使用模式对探讨物种进化、提高外源基因表达量等具重要意义。降香黄檀(*Dalbergia odorifera*)是蝶形花科(Papilionaceae)黄檀属(*Dalbergia*)常绿半落叶乔木，其心材具极高的经济和药用价值(梁远楠等, 2019; 王玥琳等, 2019; 张丽佳等, 2019)，本研究已完成对降香黄檀叶绿体基因组的高通量测序，通过分析降香黄檀叶绿体基因组蛋白编码区(coding DNA sequence, CDS)序列的碱基组成、中性绘图、ENC-plot 及 PR2-plot 等方法推断影响降香黄檀叶绿体密码子偏好性的主要因素，确定降香黄檀叶绿体基因组的最优密码子。本研究通过对降香黄檀叶绿体基因组密码子使用模式和影响密码子使用偏性的因素进行分析，确定降香黄檀叶绿体基因组的密码子偏好性及最优密码子，以期为降香黄檀叶绿体基因组的应用和研究提供科学的参考依据。

1 材料与方法

1.1 材料

降香黄檀叶片采自云南省林业科学院热带林业研究所。将采集的新鲜叶片样品保存在干冰环境中送至生工生物工程(上海)股份有限公司进行叶绿体基因组的测序，并将结果提交至 NCBI(登录号：KC427274)，通过 GenBank Feature Extractor(http://www.bioinformatics.org/sms2/genbank_feat.html)、ORF Finder(https://www.ncbi.nlm.nih.gov/orffinder/)、DNA man 等软件分析降香黄檀叶绿体基因组，根据各基因的注释结果获得 83 条 CDS 序列，剔除序列长度小于 300 bp、重复基因和 CDS 内部存在终止密码子的序列后得到 52 条符合条件的 CDS 用于分析。

1.2 方法

1.2.1 密码子组成分析

将 52 条符合条件的 CDS 整合到一个.fasta 文件中，应用 Codon W 1.4.2 软件分析获得各 CDS 的 ENC(effective number of codon, 有效密码子数)、同义密码子相对使用度(RSCU)，应用在线软件 CUSP (http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp)分析获得密码子第 1、2、3 位碱基的 GC 含量(分别为 GC_1、GC_2、GC_3)及 3 位碱基的 GC 平均含量(GC_{all})等参数，应用 SPSS 和 Excel 等数理的统计分析软件对结果进行分析。

ENC(effective number of codon, 有效密码子数)是衡量同义密码子使用偏度的重要指标，ENC 的取值范围为 20 ~ 61，ENC 值的大小能够反映密码子偏性的强弱，当 ENC 值为 20 时，代表同义密码子完全处于偏倚状态；当 ENC 值为 61 时，代表同义密码子完全没有偏倚；

ENC 值从小到大表示偏倚性由强变弱，通常可以 ENC 值 45 作为区分偏倚性强弱的标准(秦政等, 2018)。RSCU 是指某一密码子实际使用频率与无使用偏性时理论频率的比值，无偏性时，RSCU 为 1；RSCU 小于 1 则代表该密码子的实际使用频率低于其他同义密码子的使用频率，反之实际频率高于其他同义密码子的使用频率(晁岳恩等, 2012)。

应用 SPSS 软件分析密码子不同位置 GC 碱基组成 GC_1、GC_2、GC_3、GC_{all}，密码子数(N)与 ENC 等的相关关系，进而判断各因子对密码子偏好性的影响。

1.2.2 中性绘图分析

简并密码子第 3 位碱基通常所发生的为同义突变，而在密码子第 1 位、第 2 位上发生的碱基突变通常会改变基因的功能或活性；即在不存在外界压力时，密码子 3 个位置的碱基含量组成应该无差异；而在存在一定选择压力情况下时，密码子 3 个位置上的碱基组成是存在差异的(Sueoka, 2001)。在以 GC_1 和 GC_2 的平均值 GC_{12} 和 GC_3 分为纵、横坐标的中性绘图中，每一个散点代表一个基因。 如果中性图中的所有基因均沿对角线分布，即 GC_{12} 和 GC_3 的变异基本一致，密码子 3 个位置上的碱基组成无明显差异，受选择压力较弱，而受突变影响较大；回归系数(对角线斜率)是衡量中性程度的指标之一，若回归曲线斜率极小，GC_{12} 和 GC_3 的变异的相关性同样很低，说明影响密码子偏好性的主要影响因素为选择效应(Sueoka, 2001)。同时通过分析密码子不同位置碱基组成的相关性可分析密码子偏好性受突变或选择的影响，即当 GC_{12} 与 GC_3 呈显著相关时，说明密码子 3 个位置的碱基组成无明显差异，偏好性主要受突变影响； 当 GC_{12} 与 GC_3 呈不显著相关时，回归系数趋近于 0，说明密码子的前两位碱基与第 3 位碱基组成不同，基因组中的 GC 含量较为保守，密码子的使用偏性主要受选择影响(杨国锋等, 2015)。

1.2.3 ENC-plot 分析

ENC-plot 分析可探讨 ENC 和 GC_{3S} 分布的关系，是一种通过对基因数据的密码子偏好性情况的图像可视化的有效方式，其中的标准曲线代表无选择压力存在时，密码子偏好性完全由突变决定，即完全由核酸序列组成决定密码子偏好性(Wright, 1990)。

ENC-plot 绘图分析含散点图和标准曲线，散点图则以 ENC 为纵坐标，GC_3 为横坐标，标准曲线公式为 $ENC=2+GC_3+29/(GC_3^2+(1-GC_3)^2)$；具体判断标准为图中散点与标准曲线的距离，散点与标准曲线的距离近则说明密码子偏好性主要由碱基组成决定，受翻译选择的影响较为微弱；距离远则说明密码子的 ENC 值偏低，与基因表达水平存在较强的显著性相关关系，密码子偏好性较强(尚明照等, 2011；王鹏良等, 2018)。然而，ENC-plot 绘图分析不足以准确区分中性突变和选择压力的影响程度，若当选择出现在密码子第 3 位碱基上时，需要结合

ENC 比值频数对差异进行量化分析 (Sueoka, 2001)。

1.2.4 PR2-plot 分析

分析各密码子第 3 位上的 A、T、C、G 的含量，以 $A_3/(A_3+T_3)$为纵坐标，以 $G_3/(G_3+C_3)$为横坐标进行 PR2 偏倚分析(PR2-bias plot analysis)作图，用平面图显示各基因的碱基组成，其中心点代表无偏性使用时的密码子状态，即 A=T 且 C=G，其余点与中心点的矢量距离则代表其偏倚程度和方向(杨国锋等, 2015)。

1.2.5 最优密码子的确定

以降香黄檀叶绿体各基因的 ENC 作为偏好性参考标准，从两端各选择 10 %的基因构建高低偏性库，将两库间Δ RSCU≥ 0.08 的密码子定为高表达优越密码子；将 RSCU 值大于 1 的密码子确定为高频密码子(尚明照等, 2011)。将同时满足高频率密码子和高表达优越密码子确定为最优密码子。

2 结果与分析

2.1 密码子碱基组成

用在线软件 CUSP 分析降香黄檀 52 条 CDS 的碱基组成，用 Codon W 1.4.2 分析其 ENC 值(表 1)，所有 CDS 密码子的平均 GC 含量为 37.60%，第 1 位 GC 含量为 46.01%，第 2 位为 38.98%，第 3 位为 27.80%，GC 含量在密码子的不同位置分布频率不同，由高到低依次为第 1 位>第 2 位>第 3 位，说明在降香黄檀中，叶绿体基因组密码子末位碱基以 A/U(T)为主，与植物叶绿体基因中 A/U(T)含量较高的特征相符。表示偏倚强弱的 ENC 值的范围为 37.66 ~ 54.43，平均值为 46.76；52 个 CDS 的密码子 ENC 值>45 的有 36 个，说明其降香黄檀大部分基因编码序列的密码子偏性较弱。

密码子不同位置碱基的 GC 含量、密码子数(N)与 ENC 值间的相关性分析(表 2)显示，GC_{all} 和 GC_1、GC_2、GC_3 的相关性均达到极显著水平，GC_1 和 GC_2 达极显著相关， GC_3 与 GC_1、GC_2 均未达到显著相关，说明密码子的第 1 位和第 2 位碱基组成相似，与第 3 位碱基组成存在差异。ENC 与 GC_1 呈显著相关，与 GC_2 相关性不显著，与 GC_3 呈极显著相关，说明 ENC 与密码子第 3 位碱基组成密切相关。ENC 与密码子数(N)呈显著相关，说明基因编码序列长度对密码子使用偏性具一定影响。

各编码氨基酸的 RSCU(表 3)显示，RSCU >1 的密码子以 A 和 U 结尾的频率较高，其中 16 个以 U 结尾、12 个以 A 结尾、1 个以 G 结尾，说明降香黄檀叶绿体基因组偏爱以 A 和 U 结尾；而以 C 和 G 结尾的密码子为非偏爱密码子。

表 1 降香黄檀叶绿体 CDS 密码子各位置的 GC 含量

Table 1 GC proportion in different position of each CDS from the chloroplast genome of *Dalbergia oleifera*

基因 Genes	GC_1	GC_2	GC_3	GC_{all}	ENC	基因 Genes	GC_1	GC_2	GC_3	GC_{all}	ENC
accD	38.84	34.86	29.08	34.26	45.54	*psbA*	49.72	43.22	31.64	41.53	42.35
atpA	54.60	40.12	27.98	40.90	47.67	*psbB*	55.21	46.56	30.45	44.07	49.03
atpB	56.91	40.88	28.46	42.08	48.09	*psbC*	53.38	45.78	32.28	43.81	47.64
atpE	47.76	38.81	27.61	38.06	47.90	*psbD*	51.98	43.22	31.64	42.28	44.56
atpF	45.65	32.07	30.98	36.23	49.66	*rbcL*	57.14	43.49	29.83	43.49	47.75
atpI	47.98	35.08	25.00	36.02	42.30	*rpl14*	50.41	37.4	28.46	38.75	51.06
ccsA	29.63	36.73	25.93	30.76	41.62	*rpl16*	50.00	52.94	29.41	44.12	42.05
cemA	38.26	29.13	31.74	33.04	53.55	*rpl2*	50.92	48.35	31.5	43.59	54.43
clpP	58.38	36.55	30.46	41.79	51.97	*rpl20*	35.00	37.50	25.00	32.50	48.31
matK	36.24	29.26	28.68	31.40	49.51	*rpoA*	43.98	30.12	25.3	33.13	44.62
ndhA	40.66	37.91	21.15	33.24	43.48	*rpoB*	49.58	37.35	28.85	38.59	49.33
ndhB	42.19	39.15	31.85	37.73	49.36	*rpoC1*	49.19	37.19	25.18	37.19	49.00
ndhC	45.45	33.88	24.79	34.71	44.88	*rpoC2*	41.47	35.19	26.16	34.27	46.81
ndhD	37.70	37.96	30.37	35.34	47.25	*rps11*	52.52	54.68	24.46	43.88	51.52
ndhE	39.22	33.33	26.47	33.01	43.95	*rps12*	52.42	48.39	32.26	44.35	42.85
ndhF	35.98	33.07	23.54	30.86	42.76	*rps14*	43.56	47.52	24.75	38.61	39.85
ndhG	44.63	35.59	23.16	34.46	44.13	*rps18*	36.54	40.38	25.00	33.97	37.66
ndhH	51.27	35.53	25.63	37.48	49.12	*rps2*	40.51	40.93	27.85	36.43	46.71
ndhI	41.57	36.14	23.49	33.73	47.31	*rps3*	42.92	33.79	21.00	32.57	47.49
ndhJ	49.69	37.11	27.04	37.95	48.08	*rps4*	50.99	36.14	24.75	37.29	45.40
ndhK	45.79	44.39	27.57	39.25	47.67	*rps7*	52.56	44.87	23.72	40.38	46.53
petA	53.58	36.45	29.91	39.98	47.00	*rps8*	37.78	40.74	26.67	35.06	41.20
petB	46.76	41.67	30.09	39.51	42.20	*ycf1*	33.56	28.21	25.09	28.95	47.45
petD	50.31	39.13	21.12	36.85	38.83	*ycf2*	41.84	34.14	36.46	37.48	52.54
psaA	51.66	43.28	32.49	42.48	51.63	*ycf3*	47.93	38.46	30.18	38.86	51.50
psaB	48.57	43.13	30.34	40.68	48.28	*ycf4*	42.19	39.06	32.81	38.02	50.02
平均数 Average	46.01	38.98	27.80	37.60	46.76						

注：GC_{all} 表示密码子各位置的平均数。

Notes: GC_{all} represents the average of different position of codons.

表 2 密码子各位置 GC 含量、数量与 ENC 值的相关性分析

Table 2 Correlation analysis of GC content of different codon position, numbers and ENC value

变量 Variation	GC_1	GC_2	GC_3	GC_{all}	ENC
GC_2	0.504**				
GC_3	0.249	0.201			
GC_{all}	0.861**	0.805**	0.515**		
ENC	0.251*	-0.088	0.466**	0.230	
密码子数(N) Codon numbers(N)	-0.156	-0.277*	0.262*	-0.142	0.270*

注：**在 0.01 水平上显著相关；*在 0.05 水平上显著相关。

Notes: ** means significant correlation at $P< 0.01$; * means significant correlation at $P< 0.05$.

表 3 降香黄檀各氨基酸的 RSCU 分析

Table 3 RSCU analysis of protein coding region in *Dalbergia oleifera*

AA	密码子 Codons	数目 Number	RSCU	AA	密码子 Codons	数目 Number	RSCU	AA	密码子 Codons	数目 Number	RSCU
Phe	UUU	843	1.36	Pro	CCU	324	1.50	Lys	AAA	938	1.55
	UUC	393	0.64		CCC	245	0.99		AAG	271	0.45
Leu	UUA	738	1.99		CCA	246	1.14	Asp	GAU	694	1.65
	UUG	492	1.32		CCG	117	0.54		GAC	147	0.35
	CUU	447	1.20	Thr	ACU	430	1.63	Glu	GAA	847	1.49
	CUC	134	0.36		ACC	197	0.75		GAG	293	0.51
	CUA	286	0.77		ACA	331	1.25	Cys	UGU	176	1.46
	CUG	133	0.36		ACG	99	0.37		UGC	65	0.54
Ile	AUU	932	1.46	Ala	GCU	524	1.80	Arg	CGU	297	1.41
	AUC	350	0.55		GCC	192	0.66		CGC	80	0.38
	AUA	633	0.99		GCA	332	1.14		CGA	290	1.37
Val	GUU	414	1.48		GCG	119	0.41		CGG	82	0.39
	GUC	124	0.44	Tyr	UAU	656	1.65		AGA	391	1.85
	GUA	429	1.53		UAC	141	0.35		AGG	127	0.60
	GUG	155	0.55	His	CAU	388	1.56	Ser	UCU	468	1.79
Gly	GGU	506	1.36		CAC	110	0.44		UCC	245	0.94
	GGC	148	0.40	Gln	CAA	613	1.59		UCA	301	1.15
	GGA	589	1.58		CAG	160	0.41		UCG	152	0.58
	GGG	246	0.66	Asn	AAU	838	1.57		AGU	303	1.16
					AAC	229	0.43		AGC	127	0.38

注：下划线表示最优密码子。

Notes: The codon having underline represents the optimal codon.

2.2 中性绘图分析

降香黄檀叶绿体基因组各基因的中性绘图分析(图 1)，GC_{12} 的取值范围稍大在 0.309 ~ 0.536 之间，GC_3 的取值范围很小在 0.210 ~ 0.365 之间，且各基因均落在对角线上方；GC_{12} 与 GC_3 的相关系数 0.249 6，相关性不显著，且回归系数(即趋势线的斜率)为 0.393 8，即在降香黄檀叶绿体基因组中性绘图分析中，GC_{12} 和 GC_3 两个变量间的相关性很弱，说明密码子第 1 位、第 2 位和第 3 位碱基组成存在差异，即降香黄檀叶绿体基因组中 GC 含量高度保守，且密码子第 3 位的 GC 含量较低，其密码子偏好性更多地受选择的影响。

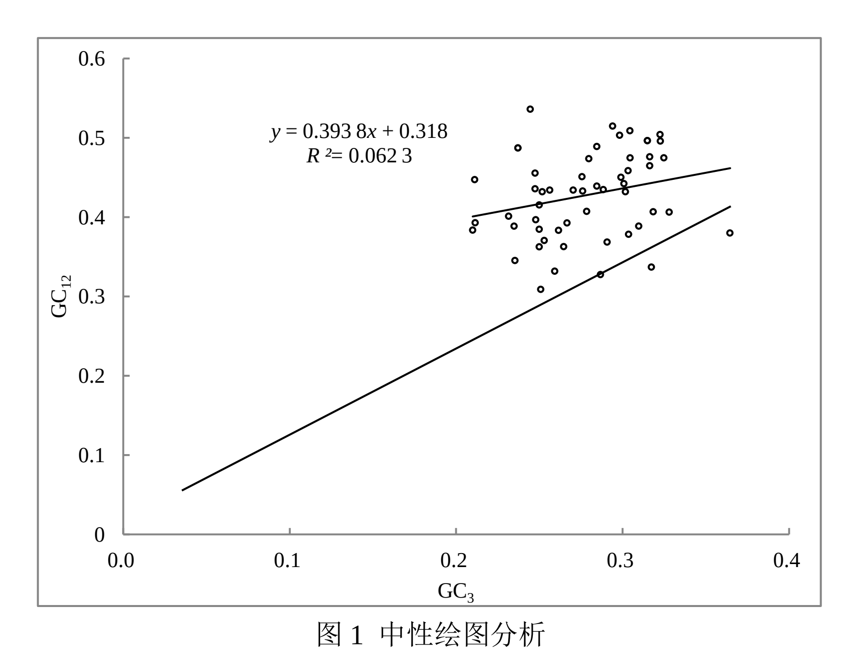


图 1 中性绘图分析

Fig.1 Neutrality plot analysis

2.3 ENC-plot 绘图分析

ENC-plot 绘图分析是以 ENC 比值[(预期 ENC 值-实际 ENC 值)/预期 ENC 值]和 ENC 比值频数来判断影响基因的主要因素，与标准曲线的距离近的基因数量较多，则其偏好性主要受突变的影响；反之选择为主要影响影响。分析结果(图 2)显示 ENC 比值频数表(表 4)中分布在-0.05 ~ 0.05 区间的基因有 14 个，即有 14 个基因与预期 ENC 值较接近，而分布在-0.05 ~ 0.05 区间之外的基因有 38 个，这 38 个基因与预期 ENC 值较远，与标准曲线的距离较远。本研究中的降香黄檀叶绿体基因组中的大多数基因与标准曲线距离较远，说明降香黄檀叶绿体基因组密码子偏好性更多地受选择的影响，而受突变的影响较弱。

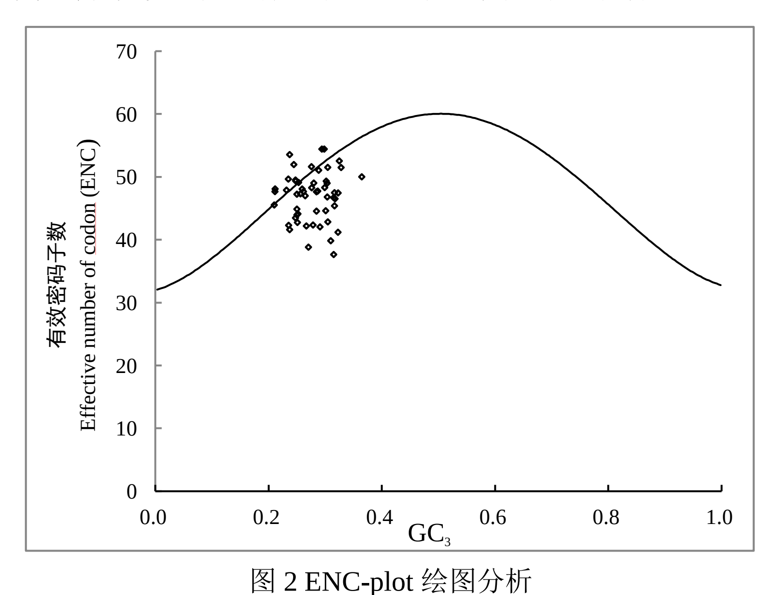


图 2 ENC-plot 绘图分析

Fig.2 Analysis of ENC and GC3 relationship

表 4 ENC 比值频数分布表

Table 4 Distribution of ENC ration

组段 Class range	组中值 Class mid value	频数 Frequency number	频率 Frequency
-0.15 ~ -0.05	-0.10	5	0.10
-0.05 ~ 0.05	0	14	0.27
0.05 ~ 0.15	0.10	25	0.48
0.15 ~ 0.25	0.20	6	0.12
0.25 ~ 0.35	0.30	2	0.04
合计 Total		52	1

2.4 PR2-plot 绘图分析

通过 PR2-plot 绘图进一步分析降香黄檀叶绿体基因组密码子偏好性的影响(图 3)，PR2 平面图中 4 个区域的散点分布不均匀，大部分基因位于处于平面图的下半部，尤其是右下方的基因数量最多，说明在碱基的使用频率方面，U＞A、G＞C。如果密码子偏好性完全受到突变的影响，则 4 个碱基的使用频率相当，因此，降香黄檀叶绿体基因组密码子偏好性同时受到突变和选择的影响。

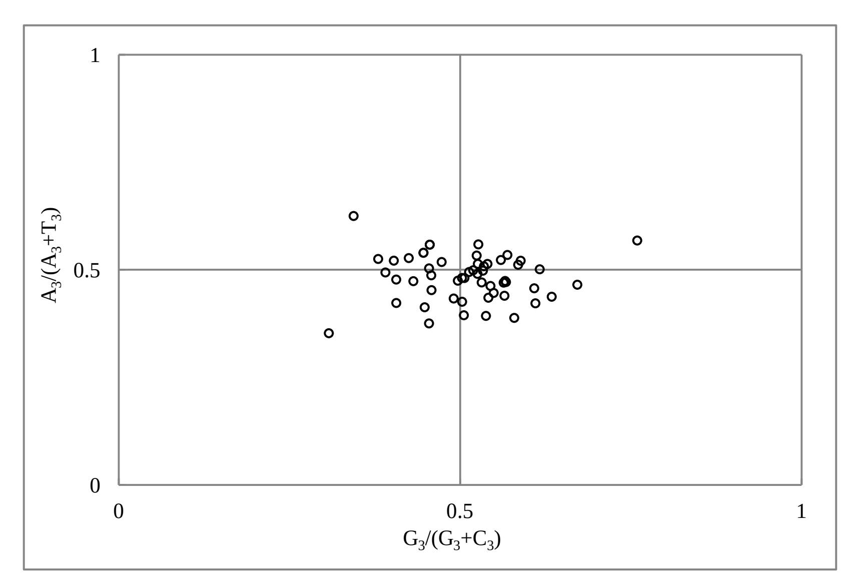


图 3 PR2-plot 绘图分析

Fig.3 Analysis of PR2 bias plot

2.5 最优密码子确定

将降香黄檀叶绿体基因组中的 52 条蛋白编码基因作为一个整体在 Codon W 1.4.2 软件上运行，通过构建高表达基因和低表达基因库，分别计算这两个基因库中的 RSCU 值，结果显示(表 5)，Δ RSCU≥0.08 的密码子，分别为 UUU(TTT)、UUA(TTA)等 23 个密码子为高表达的优越密码子(表 5 中用*表示)，其中 11 个以 A 结尾，8 个以 U 结尾，3 个以 C 结尾，1 个为 G 结尾；Δ RSCU≥0.3(表 5 中用**表示)的密码子有 11 个；Δ RSCU≥0.5(表 5 中用

***表示)的密码子分别为 UUU、UUA。以同时满足高频率密码子和高表达优越密码子作为最优密码子，分别为 UUU、UUA、AUA、GUA、UCA、AGU、CCA、UAU、GCA、CAU、CAA、AAU、AAA、GAA、UGU、CGA、AGA、GGU 和 GGA 等 19 个，其中 12 个以 A 结尾，7 个以 U 结尾。

表 5 降香黄檀叶绿体基因组最优密码子的确定

Table 5 Preferred codons in chloroplast genome of *Dalbergia odorifera*

氨基酸 AA	密码子 Codon	高表达基因 High expression gene 数目 Number	高表达基因 High expression gene RSCU	低表达基因 Low expression gene 数目 Number	低表达基因 Low expression gene RSCU	ΔRSCU	氨基酸 AA	密码子 Codon	高表达基因 High expression gene 数目 Number	高表达基因 High expression gene RSCU	低表达基因 Low expression gene 数目 Number	低表达基因 Low expression gene RSCU	ΔRSCU
Phe	UUU***	43	1.69	132	1.07	0.62	Ala	GCU	17	1.39	70	1.62	-0.23
	UUC	8	0.31	114	0.93	-0.62		GCC	6	0.49	31	0.72	-0.23
Leu	UUA***	56	3.23	85	1.25	1.98		GCA**	21	1.71	54	1.25	0.46
	UUG	20	1.15	96	1.41	-0.26		GCG	5	0.41	18	0.42	-0.01
	CUU	21	1.21	94	1.38	-0.17	His	CAU*	14	1.65	97	1.56	0.09
	CUC	0	0.00	36	0.53	-0.53		CAC	3	0.35	27	0.44	-0.09
	CUA	5	0.29	63	0.92	-0.63	Gln	CAA**	22	1.76	102	1.45	0.31
	CUG	2	0.12	35	0.51	-0.39		CAG	3	0.24	39	0.55	-0.31
Ile	AUU	42	1.34	134	1.30	0.04	Asn	AAU*	39	1.70	155	1.49	0.21
	AUC	16	0.51	71	0.69	-0.18		AAC	7	0.30	53	0.51	-0.21
	AUA*	36	1.15	105	1.02	0.13	Lys	AAA**	49	1.75	144	1.38	0.37
Val	GUU	15	1.33	56	1.27	0.06		AAG	7	0.25	65	0.62	-0.37
	GUC	7	0.62	29	0.66	-0.04	Asp	GAU	14	1.40	160	1.65	-0.25
	GUA**	20	1.78	65	1.48	0.30		GAC*	6	0.60	34	0.35	0.25
	GUG	3	0.27	26	0.59	-0.32	Glu	GAA*	32	1.52	123	1.28	0.24
Ser	UCU	21	1.62	89	1.61	0.01		GAG	10	0.48	69	0.72	-0.24
	UCC	6	0.46	63	1.14	-0.68	Cys	UGU**	8	1.78	33	1.43	0.35
	UCA*	19	1.46	66	1.20	0.26		UGC	1	0.22	13	0.57	-0.35
	UCG**	14	1.08	36	0.65	0.43	Arg	CGU*	14	1.06	39	0.98	0.08
	AGU*	17	1.31	58	1.05	0.26		CGC	3	0.23	16	0.40	-0.17
	AGC	1	0.08	19	0.34	-0.26		CGA*	21	1.59	52	1.30	0.29
Pro	CCU	14	1.30	57	1.42	-0.12		CGG	5	0.38	24	0.60	-0.22
	CCC*	10	0.93	33	0.82	0.11		AGA**	29	2.20	70	1.75	0.45
	CCA*	15	1.40	46	1.15	0.25		AGG	7	0.53	39	0.98	-0.45
	CCG	4	0.37	24	0.60	-0.23	Gly	GGU*	24	1.45	79	1.32	0.13
Thr	ACU	18	1.33	62	1.39	-0.06		GGC	9	0.55	25	1.42	-0.87
	ACC**	15	1.11	35	0.79	0.32		GGA	25	1.52	89	1.48	0.04
	ACA	14	1.04	61	1.37	-0.33		GGG	8	0.48	47	0.78	-0.30
	ACG	7	0.52	20	0.45	0.07							
Tyr	UAU**	26	1.86	100	1.54	0.32							
	UAC	2	0.14	30	0.46	-0.32							

*表示Δ RSCU≥0.08, **表示Δ RSCU ≥0.3, ***表示Δ RSCU ≥0.5。

* meanΔ RSCU ≥0.08, ** meanΔ RSCU ≥0.3, *** meanΔ RSCU ≥0.5.

3 讨论与结论

在生物体中，密码子在核酸和蛋白质的翻译方面扮演着重要作用；植物中不同密码子的使用频率存在差异，这种密码子使用偏好性是物种和基因长期进化和对环境的适应过程中形成的，是由于多种因子共同作用的结果，其中突变和自然选择是该现象形成的重要影响因素(Romero et al., 2000; Xu et al., 2011)。叶绿体是植物进行光合作用的细胞器，亦含有相对独立的母系遗传基因组信息，故叶绿体基因组在揭示物种进化、不同物种间亲缘关系、物种鉴定等方面具重要价值；同时叶绿体基因工程因其可高效表达、安全等特点已成为植物基因工程的研究热点(Wright, 1990; Duret, 2000)。因此，植物叶绿体基因组密码子使用偏性的研究能够揭示物种基因组进化关系及主要影响因素。

中性进化理论认为碱基突变和自然选择对密码子第 3 位碱基变化的影响是中性或近中性的(Sharp et al., 1993)，即通常密码子的第 1 位、第 2 位碱基的改变会造成编码氨基酸的改变，而第 3 位碱基改变则对氨基酸编码无影响，可以认为密码子使用偏性在某种程度上是对偏好密码子使用与非偏好密码子保留间的一种平衡，是进化过程中的一种自我保护机制；同时因密码子第 3 位碱基具有的兼并性及较小的选择压力、GC_3 含量与密码子使用偏性的显著相关性等因素，通常将 GC_3 作为密码子使用模式分析的重要依据(Gu et al., 2004; Ingvarsson, 2007)。本研究中降香黄檀叶绿体基因组密码子的 GC_3 含量远低于前两位，与蒺藜苜蓿(*Medicago truncatula*)(杨国锋等, 2015)、紫花苜蓿(*Medicago sativa*)(陶晓丽, 2017)的密码子 3 个位置的 GC 含量进行比较，即紫花苜蓿 GC_1(45.24 %) > GC_2(37.30 %) > GC_3(28.97 %)，蒺藜苜蓿 GC_1(45.5 %) > GC_2(36.8%) > GC_3(26.9 %)，3 个物种叶绿体基因组在密码子 3 个位置上的 GC 含量趋势一致，但具体数据存在一定的差异。中性绘图分析显示，密码子第 1 位和第 2 位与第 3 位碱基组成存在显著差异，通过其 GC 含量高度保守，其密码子偏好性主要受选择影响；与蒺藜苜蓿一致(杨国锋等, 2015)；同时结合 ENC-plot 和 PR2-plot 等分析方式发现降香黄檀叶绿体基因组密码子的偏好性受多种因素综合影响，主要影响因素为选择；这个结论与其同科植物蒺藜苜蓿(*Medicago truncatula*)(杨国锋等, 2015)一致。降香黄檀叶绿体基因组密码子偏好以 AT 结尾，且其最优密码子为 UUU、UUA，与大多数高等植物的最优密码子 NNA、NNU 的模式一致(尚明照等, 2011)。这种密码子使用模式可能由于叶绿体基因组中含有丰富的 AT 碱基，且其密码子使用模式存在显著差异，而在进化和亲缘关系较近的植物通常具相似的密码子使用模式，叶绿体基因组密码子的偏好性在进化关系上较为保守。本研究中降香黄檀叶绿体基因组密码子偏好性主要受突变的影响，同时与其他因素一起共同作

用于其密码子使用模式，确定了 19 个最优密码子，且均为 NNA 和 NNU 模式。降香黄檀的分布区较为零散，仅分布海南全岛，及广东、广西和福建省的少数地区(张丽佳等, 2019)，但降香黄檀叶绿体基因组密码子的偏好性与其他蝶形花科植物基本一致，未表现出特异性，说明蝶形花科植物至少在密码子偏好性方面亦较为保守。本研究为以后降香黄檀通过外源基因密码子改造的异源表达、叶绿体基因工程和遗传多样性分析提供了科学的参考依据。

参考文献:

CHAO YE, CHANG Y, WANG MF, et al., 2012. Codon usage bias and cluster analysis on chloroplastic genes from seven crop species[J]. Acta Agric Bor Sin, 27 (4):60-64.[晁岳恩, 常阳, 王美芳, 等, 2012. 7 种作物叶绿体基因的密码子偏好性及聚类分析[J]. 华北农学报, 27 (4):60-64.]

DURET L, 2000. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes[J]. Trends Genet, 16 (7):287-289.

DURET L, 2002. Evolution of synonymous codon usage in metazoans[J]. Curr Opin Genet Dev, 12 (6):640-649.

GU W, ZHOU T, MA J, et al., 2004. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*[J]. Biosystems, 73 (2):89-97.

INGVARSSON PK, 2007. Gene expression and protein length influence codon usage and rates of sequence evolution in *Populus tremula*[J]. Mol Biol Evol, 24 (3):836-844.

LIANG YN, CHEN SL, ZHANG LJ, et al., 2019. Early growth evaluation of 10 *Dalbergia odorifera* families in Zhaoqing[J]. For Environ Sci, 35 (2):106-110.[梁远楠, 陈水莲, 张丽君, 等, 2019. 10 个降香黄檀家系在肇庆地区的早期生长评价[J]. 林业与环境科学, 35 (2):106-110.]

NIU Y, XU Q, WANG YD, et al., 2018. An analysis on codon usage bias of chloroplast genome of *Rosa odorata* var. *gigantea*[J]. J NW Fore Univ, 33 (3):123-130.[牛元, 徐琼, 王嵛德, 等, 2018. 大花香水月季叶绿体基因组密码子使用偏性分析[J]. 西北林学院学报, 33 (3):123-130.]

QIN Z, ZHENG YJ, GUI LJ, et al., 2018. Codon usage bias analysis of chloroplast genome of camphora tree(*Cinnamomum camphora*)[J]. Guihaia, 38 (10):1346-1355.[秦政, 郑永杰, 桂丽静, 等, 2018. 樟树叶绿体基因组密码子偏好性分析[J]. 广西植物, 38 (10):1346-1355.]

ROMERO H, ZAVALA A, MUSTO H, 2000. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces[J]. Nucl Acid Res, 28 (10):2084-2090.

SHANG MZ, LIU F, HUA JP, et al., 2011. Analysis on codon usage of chloroplast genome of *Gossypium hirsutum*[J]. Sci Agric Sin, 44 (2):245-253.[尚明照, 刘方, 华金平, 等, 2011. 陆地棉叶绿体基因组密码子使用偏性的分析[J]. 中国农业科学, 44 (2):245-253.]

SHARP PM, STENICO M, PEDEN JF, et al., 1993. Codon usage: Mutational bias, translation selection, or both? [J] Biochem Soc Trans, 21(4): 835-841.

SUEOKA N, 2001. Near homogeneity of PR2-bias fingerprints in the human genome and their implications in phylogenetic analyses[J]. J Mol Evol, 53: 469-476.

TAO XL, 2017. The study on the complete chloroplast genome of *Medicago sativa* and *Vicia*

sativa[D]. Lanzhou: Lanzhou University: 22-31.[陶晓丽, 2017. 紫花苜蓿和箭筈豌豆叶绿体全基因组研究[D]. 兰州: 兰州大学: 22-31]

WANG PL, YANG LP, WU HY, et al., 2018. Codon preference of chloroplast genome in *Camellia oleifera*[J]. Guihaia, 38 (2):135-144.[王鹏良, 杨利平, 吴红英, 等, 2018. 普通油茶叶绿体基因组密码子偏好性分析[J]. 广西植物, 38 (2):135-144.]

WANG YL, XU DP, YANG ZJ, et al., 2019. Effects of pruning and ethylene on photosynthetic system characteristics in *Dalbergia odorifera*[J]. Mol Plant Breed, 17 (7):2392-2398.[王玥琳, 徐大平, 杨曾奖, 等, 2019. 修枝和乙烯对降香黄檀光合系统特性影响[J]. 分子植物育种, 17 (7):2392-2398.]

WU XM, WU SF, REN DM, et al., 2007. The analysis method and progress in the study of codon bias[J]. Hereditas, 29 (4):420-426.

WRIGHT F, 1990. The 'effective number of codons' used in a gene[J]. Gene, 87 (1):23-29.

XING CB, CAO L, ZHOU M, et al., 2013. Analysis on codon usage of chloroplast genome of *Eleutherococcus senticosus*[J]. Chin J Chin Mat Med, 38 (5):661-665.[邢朝斌, 曹蕾, 周秘, 等, 2013. 刺五加叶绿体基因组密码子的用法分析[J]. 中国中药杂志, 38 (5):661-665.]

XU C, CAI X, CHEN Q, 2011. Factors affecting synonymous codon usage bias in chloroplast genome of oncidium gower ramsey[J]. Evol Bioinform, 7(7):271-278.

YANG GF, SU KL, ZHAO YR, et al., 2015. Analysis of codon usage in the chloroplast genome of *Medicago truncatula*[J]. Acta Pratac Sin, 24 (12):171-179.[杨国锋, 苏昆龙, 赵怡然, 等, 2015. 蒺藜苜蓿叶绿体密码子偏好性分析[J]. 草业学报, 24 (12):171-179.]

ZHANG LJ, PU YJ, LI DD, et al., 2019. Effects of potassium silicate at different concentrations on the plant growth and physiological characteristics of *Dalbergia odorifera* T. Chen seedlings[J]. Nat Sci J Hainan Univ, 37 (1):6-13.[张丽佳, 蒲玉瑾, 李大东, 等, 2019. 不同浓度硅酸钾对降香黄檀幼苗生长及其生理特征的影响[J]. 海南大学学报(自然科学版), 37 (1):6-13.]

ZHANG JW, JIANG ZM, SU H, et al., 2019. The complete chloroplast genome sequence of the endangered species *Syringe pinnatifolia* (Oleaceae)[J]. Nor J Bot, 37(5). DOI: 10.1111/njb.02201.

ZHOU M, LONG W, LI X, 2008a. Patterns of synonymous codon usage bias in chloroplast genomes of seed plants[J]. For Stud Chin, 10 (4):235-242.

ZHOU M, LONG W, LI X, 2008b. Analysis of synonymous codon usage in chloroplast genome of *Populas alba*[J]. J For Res, 19 (4):293-297.